

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ТРАНСФЕРНОГО НАВЧАННЯ ЗА ДОПОМОГОЮ ДИНАМІЧНОГО ВИЗНАЧЕННЯ КІЛЬКОСТІ ШАРІВ ДЛЯ ЗАМОРОЗКИ НА ОСНОВІ ПОДІБНОСТІ КЛАСІВ

¹Державний університет «Житомирська політехніка»

Запропоновано метод, який автоматично визначає кількість шарів нейронної мережі для заморожування під час трансферного навчання. Цей підхід ґрунтується на використанні косинусної відстані між векторними представленнями класів вихідного та цільового наборів даних, що дозволяє оцінити їхню семантичну близькість і відповідно регулювати глибину заморожки шарів для подальшого перенавчання. На відміну від фіксованих стратегій, динамічний підхід забезпечує гнучкіше використання попередніх знань і зменшує потребу у надмірних обчисленнях. У роботі використано архітектуру ResNet-50 та підмножини класів набору CIFAR-10, для яких формувалися середні векторні значення та обчислювалася косинусна відстань. На основі отриманих значень динамічно визначалося, яку кількість шарів моделі доцільно заморозити та залишити незмінною, а які шари потребують перенавчання. Ефективність методу оцінювалася шляхом порівняння класичного навчання з нуля та навчання із застосуванням цього підходу трансферного навчання. Отримані результати показали, що запропонований підхід сприяє підвищенню якості узагальнення та скороченню часу тренування, демонструючи переваги використання динамічного визначення кількості шарів для заморожки. Метод може застосовуватися у задачах, де важливими є швидка адаптація, обмежені дані та економне використання ресурсів. Запропонована методика є перспективною, оскільки дозволяє ефективно поєднувати точність моделі з меншими обчислювальними витратами, забезпечуючи можливість масштабування, повторного використання певної кількості шарів та швидкої інтеграції у різні прикладні області, завдяки чому вона є цінним інструментом для подальших досліджень і практичних застосувань сьогодення.

Ключові слова: трансферне навчання, динамічне заморожування шарів, косинусна відстань, ResNet-50, CIFAR-10, скорочення ресурсів.

Вступ

Технології навчання моделей розвиваються так швидко, що тепер від моделей потрібна не лише ефективність, а й здатність швидко адаптуватися до нових завдань. Незважаючи на прогрес технологій, процес навчання моделі досі вимагає чималої кількості ресурсів та даних для навчання. Проте не завжди це можливо. Особливо це відчувається в деяких сферах — скажімо, у відеоспостереженні, автономних системах чи медичній діагностиці, де швидкість адаптації моделі до нових умов є критично важливою, а дані для навчання не завжди доступні в достатній кількості.

Трансферне навчання стало одним з ключових напрямів розв'язання цих проблем, оскільки дозволяє повторно використовувати вже навчені моделі та зменшувати потребу у повному перенавчанні. Проте класичний підхід ґрунтується на тому, що кількість шарів для заморожки визначається статично, що не враховує різний ступінь семантичної близькості між вихідним та цільовим класами. У результаті це може або обмежувати здатність моделі до адаптації, або, навпаки, призводити до зайвих витрат часу та ресурсів.

У цій роботі розглядається новий метод динамічного трансферного навчання, у якому кількість шарів для заморожування визначається автоматично на основі косинусної відстані між класами. Такий підхід дозволяє динамічно регулювати глибину перенавчання залежно від того, наскільки схожими є ознаки різних категорій.

Метою роботи є визначення ефективності динамічного трансферного навчання порівняно з традиційним підходом, а також оцінка того, наскільки нестатичне заморожування шарів може скоротити

час навчання та чи не погіршиться точність моделі. У межах дослідження демонструється як математичне обґрунтування методу, так і його експериментальна реалізація з детальним порівнянням результатів навчання моделі з нуля та моделі, перенавченої з використанням косинусної відстані та динамічним визначенням шарів для заморозки.

Проблема цього підходу полягає у тому, що навчання глибоких нейронних мереж з нуля залишається вкрай ресурсомістким процесом, що вимагає значних обчислювальних можливостей і, в деяких випадках, великих обсягів даних. У реальних умовах, де час навчання та доступні ресурси можуть бути обмежені, такий підхід не завжди є ефективним.

Динамічне визначення кількості шарів для заморозки є альтернативою до класичного методу трансферного навчання, оскільки цей підхід допомагає ефективно використовувати наявні ресурси та не потребує великої кількості даних для навчання. Проте залишається відкритим питання, наскільки цей метод є ефективним у скороченні часу на навчання та чи не погіршується при цьому ефективність самої моделі.

Саме тому дослідження ефективності такого динамічного підходу на прикладі архітектури ResNet-50 є важливим кроком для дослідження та подальшого створення адаптивних, продуктивних та економічних систем навчання, які здатні зменшувати час тренування та не погіршувати якість моделі в умовах обмежених ресурсів.

Трансферне навчання давно стало одним з ключових напрямів оптимізації глибоких нейронних мереж, а сучасні дослідження все частіше зосереджуються на автоматизації та адаптивності цього процесу. Одним з базових джерел є огляд Pan і Yang, які систематизували основні підходи трансферного навчання та класифікували механізми перенесення знань між моделями [1]. Автори підкреслили, що успішний трансфер знань значною мірою залежить від схожості класів, але при цьому не запропонували формального способу оцінювання цієї схожості в контексті вибору глибини заморожування шарів. Подібна відсутність адаптивності й стала ключовою науковою проблемою у подальших дослідженнях.

Значний внесок у розвиток динамічних підходів зробила робота Zhuang et al. (2020) [2], де автори провели масштабний огляд методів глибокого трансферного навчання та виділили перспективні напрями, зокрема автоматичне визначення релевантної кількості шарів моделі. Попри глибину аналізу, автори не запропонували конкретної методики оцінки семантичної подібності класів, через що ця робота є важливим, але непрямим джерелом для формування динамічних підходів.

Окрему увагу привертають дослідження, пов'язані з використанням метрик подібності для задач оптимізації моделей. Зокрема, Huang (2008) [3] продемонстрував ефективність косинусної відстані як універсальної міри оцінювання схожості векторних представлень у задачах кластеризації текстів. Хоча робота належить до галузі обробки текстів, її висновки напряму переносяться у простір ознак глибоких нейронних мереж: косинусна відстань відображає близькість напрямів ознак, що дозволяє адаптивно визначати, наскільки нові класи корелюють з базовими. Саме завдяки цій властивості косинусна відстань є фундаментальним інструментом для побудови динамічних стратегій трансферного навчання.

До того ж, важливими є роботи, що досліджують відмінності у перенесенні ознак між різними шарами глибоких мереж. Дослідження Yosinski et al. (2014) [4] показало, що нижні шари CNN моделюють універсальні ознаки, тоді як верхні — специфічніші. Цей висновок підтверджує необхідність адаптивного підходу: різні рівні подібності між класами повинні призводити до різних стратегій заморожування та донавчання шарів. Попри це, навіть у цій роботі не розглянуто механізм автоматичного визначення кількості шарів, які потребують перенавчання.

Власне дослідження Іванова (2024) [5] стало одним з перших кроків у формальному описанні ефективності трансферного навчання з точки зору часових витрат. У статті доведено, що статична схема заморожування шарів ResNet-50 призводить до логарифмічного зниження витрат часу порівняно з навчанням з нуля. Проте у роботі також підкреслюється обмеження — кількість заморожених шарів залишається сталою для всіх задач, незалежно від того, наскільки схожими є вихідні та цільові класи. Це створює передумови для формування нового наукового напрямку — динамічного трансферного навчання, де глибина адаптації визначається автоматично.

Отже, аналіз сучасних досліджень показує, що хоча трансферне навчання активно розвивається, більшість робіт або концентруються на загальних принципах перенесення знань, або описують ефективність окремих архітектур, але не пропонують чіткої математичної моделі адаптивного визначення глибини трансферу. Існує значна прогалина між статичними підходами та потребою у динамічних методах, які могли б визначати, скільки шарів моделі потрібно заморожувати залежно від

семантичної близькості класів. Це і є тією науковою проблемою, на вирішення якої спрямоване це дослідження.

Результати дослідження

Трансферне навчання — це підхід, за допомогою якого знання, отримані моделлю під час розв'язання одного завдання, використовуються для швидкої адаптації до іншого, потенційно відмінного за структурою. Такий механізм дозволяє значно зменшити час і обчислювальні витрати, оскільки замість повного навчання з нуля модель використовує уже сформовані ознаки та представлення [1].

Статичне заморожування фіксованої кількості шарів у процесі трансферного навчання не завжди є оптимальним, оскільки рівень подібності між вихідним і цільовим класами може як суттєво відрізнятися, так і бути дуже схожим. У випадках, коли нові класи значно віддалені від базового домену, надмірне заморожування шарів знижує якість адаптації моделі, тоді як для дуже схожих класів навпаки — перенавчання зайвих шарів призводить до нераціонального використання ресурсів. Ефективність перенесення залежить від того, наскільки особливості, сформовані мережею на вихідному наборі даних, відповідають особливостям нового завдання [6]. Це означає, що різні шари нейронних мереж суттєво відрізняються за рівнем універсальності, а отже, потреба в їх перенавчанні залежить від семантичної близькості класів [7]. Це підтверджує, що вибір кількості шарів для заморожки має виконуватися динамічно на основі реальної подібності між класами, що дозволяє досягти балансу між точністю та ефективністю навчання.

Для динамічного вибору кількості шарів, які необхідно заморозити під час трансферного навчання, ключовим є визначення ступеня подібності між класами вихідного та цільового наборів даних. Що більше схожими є класи, то більша частина базових ознак уже присутня у вихідній моделі, а отже — тим більше шарів можна заморозити без втрати якості. Для оцінювання цієї подібності використовується косинусна відстань.

Для застосування косинусної відстані необхідно подати кожен клас у вигляді векторного представлення, що відображає його ознаки у багатовимірному просторі. Найчастіше це досягається за допомогою попередньо навченої нейронної мережі, яка використовується як екстрактор ознак: вхідні зображення пропускаються через модель, після чого береться вектор з одного з проміжних шарів, зазвичай останнього перед класифікаційною головою. Для кожного класу обчислюється середнє значення таких векторів (centroid embedding), що дозволяє отримати узагальнене представлення класу у вигляді одного вектора. Такий підхід є стандартним у сучасних дослідженнях перенесення ознак [4], [6], оскільки він забезпечує стабільне та масштабоване порівняння семантичної близькості між класами за допомогою косинусної відстані.

Математично косинусна відстань обчислюється як міра збігу напрямків двох векторів. Якщо маємо два вектори A і B , то вона визначається формулою

$$\cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|}, \quad (1)$$

де $A \cdot B$ — це скалярний добуток векторів, а $\|A\|$ та $\|B\|$ — довжини векторів. Значення $\cos(A, B)$ лежить у межах від -1 до 1 : значення, близькі до 1 , означають високу подібність між класами, що дозволяє заморожувати більшу кількість шарів; тоді як нижчі значення вказують на необхідність глибшого перенавчання моделі.

Для практичної реалізації динамічного трансферного навчання важливо врахувати, що навіть за високої косинусної відстані між класами неможливо заморозити усі шари моделі, оскільки мережа повинна зберігати мінімальну здатність до адаптації. Тому косинусна відстань використовується як коефіцієнт, що визначає частку шарів, які можуть бути заморожені від максимально дозволеної кількості. За високим значенням подібності (наприклад, понад $0,8$) заморожується більша частина шарів, тоді як за низьких значень мережа вимагає глибшого перенавчання. Такий підхід узгоджується з теоретичними висновками про залежність ефективності перенесення від відстані між класами [8].

Визначити кількість шарів для заморожки можна за формулою

$$L_{freeze} = \text{round}(S \cdot L_{total}), \quad (2)$$

де S — значення косинусної відстані між вихідним і цільовим класами, що лежить у діапазоні від 0 до 1 (оскільки, якщо значення менше 0 , то немає сенсу застосовувати його у цьому підході), а

L_{total} — максимальна кількість шарів моделі, які можна заморозити без втрати здатності до адаптації.

Дослідження проведено на основі архітектури ResNet-50 та набору даних CIFAR-10, оскільки ці об'єкти є класичними еталонами у сфері комп'ютерного зору та широко застосовуються для валідації методів трансферного навчання. Архітектура ResNet-50 вибрана через її глибину, стабільність градієнтів та добре задокументовану структуру, що включає модулі різного рівня абстракції — від універсальних низькорівневих ознак до високорівневих семантичних представлень [9]. Через це вона є оптимальною моделлю для тестування методів динамічного заморожування шарів. Набір даних CIFAR-10 використано завдяки його компактності, збалансованості класів та широкому застосуванню в експериментах, що потребують багаторазового повторення або порівняння різних стратегій оптимізації [10]. До того ж, у попередніх дослідженнях продемонстровано, що ResNet-архітектури добре адаптуються до невеликих зображень з CIFAR-підмножин, що дозволяє коректно оцінити вплив косинусної відстані на ефективність динамічного трансферного навчання.

Для подальшого визначення кількості шарів, які доцільно заморозити під час трансферного навчання, необхідно попередньо оцінити ступінь семантичної близькості між класами. З цією метою кожен клас CIFAR-10 перетворено у векторне представлення за допомогою попередньо навченої моделі ResNet-50, після чого обчислено косинусну відстань між їх середніми векторами. Отримана матриця подібності дозволяє візуально та кількісно визначити, які класи мають схожі ознаки й, відповідно, потребують мінімального донавчання, а які — суттєво відрізняються, що вимагає глибшої адаптації моделі. Саме на основі цих значень надалі визначається динамічна кількість шарів для

заморозки (рис. 1).

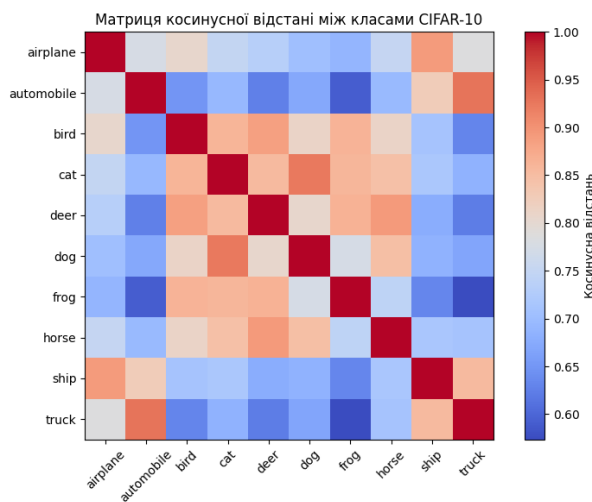


Рис. 1. Матриця косинусної подібності між класами набору даних CIFAR-10

Графік відображає ступінь семантичної близькості між середніми векторними представленнями класів. Кожний елемент матриці показує рівень схожості між двома класами: значення, близькі до 1, свідчать про високу схожість ознак, тоді як нижчі значення вказують на суттєві відмінності між класами. Така матриця використовується для візуалізації того, в яких випадках класи схожі і можуть заморожувати більшу кількість шарів.

Для наочного аналізу просторової структури ознак, що формуються попередньо навченою ResNet-50, додатково виконано проєкцію усереднених векторів класів CIFAR-10 у двовимірний простір методом головних компонент (PCA). Така візуалізація дозволяє оцінити, наскільки класи відокремлені один від одного у високовимірному просторі ознак, та виявити групи класів, що мають подібні представлення. Оскільки PCA зберігає найінформативніші напрямки варіації даних, вона слугує інструментом для інтерпретації семантичної структури ознак і допомагає зрозуміти, чому саме деякі класи демонструють вищу косинусну відстань ніж інші [9]. Отримана проєкція використовується як додатковий доказ коректності підходу динамічного визначення кількості шарів для заморожки.

На рис. 2 показано розташування усереднених векторних представлень класів CIFAR-10 після зменшення їх розмірності до двох головних компонент. Близьке розташування точок вказує на схожість ознак, які модель ResNet-50 виділяє для відповідних класів, тоді як значне розділення між точками означає відмінність у їхніх представленнях. Така візуальна інтерпре-

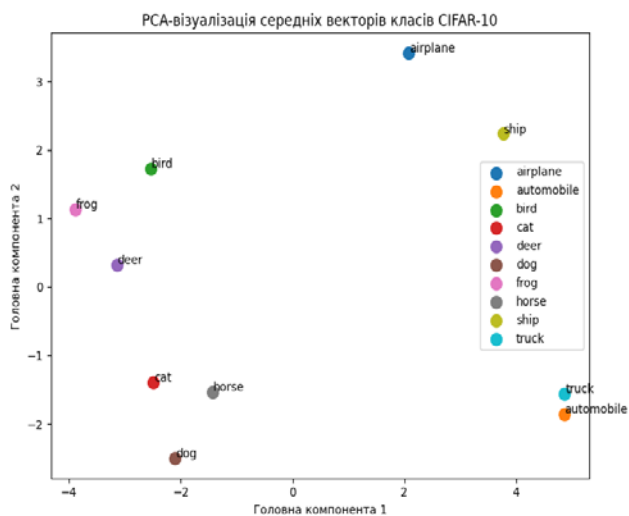


Рис. 2. PCA-візуалізація центротидів класів CIFAR-10 у двовимірному просторі

тація доповнює матрицю косинусної відстані та демонструє, які класи можуть потребувати мінімального перенавчання, а для яких необхідне глибше навчання, або ж взагалі — повний цикл навчання.

Архітектура ResNet-50 складається з п'яти основних компонентів: початкового блоку conv1 з нормалізацією та pooling-шаром, а також чотирьох груп bottleneck-блоків (layer1–layer4), які формують 49 із 50 згорткових підшарів моделі [9]. Хоча теоретично модель містить 50 тренуваних шарів, заморозити їх усі під час трансферного навчання неможливо, оскільки останні шари відповідають за формування високорівневих семантичних ознак та класифікаційних рішень. Зокрема, останні 10 шарів містять параметри, необхідні для адаптації до нового набору класів, тому їх обов'язково потрібно перенавчати. Саме ці шари відповідають за трансформацію універсальних ознак у специфічні для конкретного завдання, і їх заморожування призвело б до неможливості моделі розрізняти нові класи. Таким чином, максимально можлива кількість шарів, яку доцільно заморозити, обмежується приблизно 40, що відповідає глибині перших трьох груп блоків, які містять універсальні ознаки нижчого рівня.

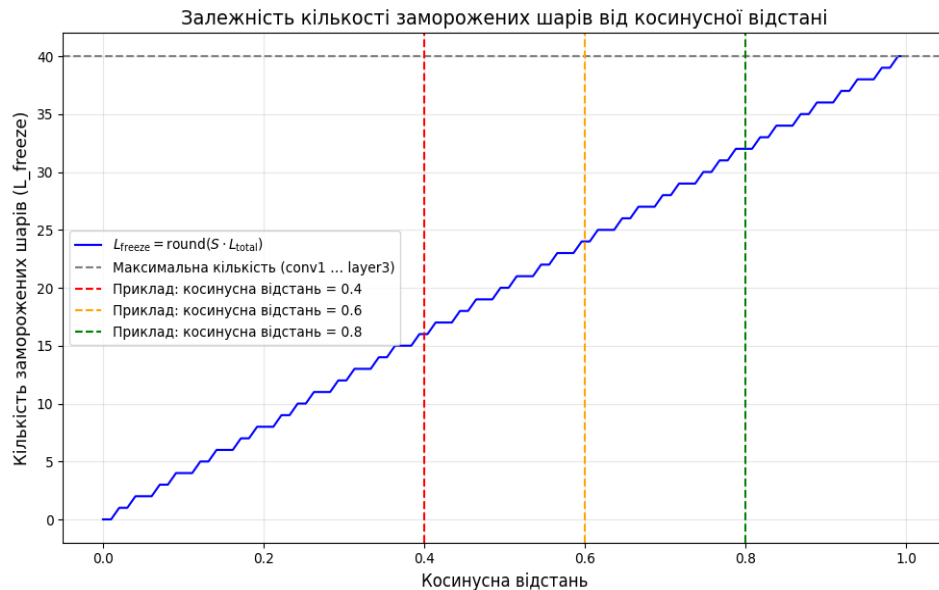


Рис. 3. Залежність кількості заморожених шарів моделі від косинусної відстані між класами

На графіку (рис. 3) показано лінійну залежність між косинусною відстанню класів та кількістю шарів ResNet-50, які доцільно заморозити під час трансферного навчання. Графік демонструє, як кількість заморожених шарів зростає разом зі збільшенням косинусної відстані між класами. Синя лінія відображає плавну лінійну залежність: починаючи з нуля, коли $S = 0$, вона зростає до максимально допустимого значення 40 шарів, яке відповідає верхній межі (сіра пунктирна лінія). Це значення узгоджується з архітектурою ResNet-50, де лише перші чотири групи блоків (conv1–layer3) містять універсальні ознаки, придатні до заморозки.

На графіку нанесено три контрольні точки, позначені кольоровими вертикальними лініями, які ілюструють застосування формули на конкретних прикладах. Червона лінія означає, що модель повинна перенавчати більшу частину параметрів, оскільки класи цільового набору даних суттєво відрізняються від вихідних і використання трансферного навчання не завжди є доцільним. Помаранчева лінія показує, що модель уже може використовувати більші частини попередньо навченої структури, і використання цього методу вже є доцільним. Зелена лінія означає, що модель зберігає більшість раніше вивчених ознак, а перенавчання підлягають лише високорівневі шари, пов'язані з адаптацією до конкретного завдання і використання трансферного навчання значного пришвидшить процес навчання моделі.

Для перевірки чи покращує цей метод процес навчання, проведено експеримент. Виконано три різних навчання моделі — навчання моделі з нуля для визначення котів, потім для визначення собак та використання цього методу для навчання моделі для визначення собак з використанням попередньо навченої моделі на котях. У кожному випадку модель навчається на однаковому обсязі даних: з набору CIFAR-10 відбирається по 800 зображень класу «кіт» та 800 зображень класу «собака», після чого формується збалансований датасет обсягом 1600 зразків. Для кожного експерименту 80 % цих зразків (1280 зображень) використовуються для навчання, а решта 20 % (320 зображень) — для

валідації. Це забезпечує коректне порівняння методів, оскільки всі моделі навчаються на тому самому обсязі даних, а різниця виникає лише у початкових вагових параметрах та у кількості заморожених шарів у випадку трансферного навчання.

Косинусна відстань (кіт-собаки): 0.9225
Кількість зображень – 800

[01] трен.втрата 0.5122 точн. 0.755 вал.втрата 0.6574 точн. 0.766
[02] трен.втрата 0.3551 точн. 0.865 вал.втрата 0.5561 точн. 0.769
[03] трен.втрата 0.1712 точн. 0.939 вал.втрата 0.8632 точн. 0.766
[KIT з нуля] найкраща вал.точність = 0.769, час = 2951.4 с
[01] трен.втрата 0.5214 точн. 0.752 вал.втрата 1.5049 точн. 0.641
[02] трен.втрата 0.3857 точн. 0.842 вал.втрата 0.9630 точн. 0.631
[03] трен.втрата 0.2240 точн. 0.916 вал.втрата 0.5126 точн. 0.803
[СОБАКА з нуля] найкраща вал.точність = 0.803, час = 2879.7 с
[01] трен.втрата 0.7261 точн. 0.616 вал.втрата 0.3251 точн. 0.816
[02] трен.втрата 0.2293 точн. 0.909 вал.втрата 0.2744 точн. 0.875
[03] трен.втрата 0.1248 точн. 0.956 вал.втрата 0.3175 точн. 0.884
[KIT → СОБАКА трансферне навчання] найкраща вал.точність = 0.884, час = 2234.6 с

Рис. 4. Результати дослідження методу динамічного трансферного навчання

Результати експериментів демонструють, що запропонований метод динамічного трансферного навчання забезпечує значно кращу узагальнювальну здатність моделі порівняно з навчанням з нуля. За умов використання 800 зображень для кожного класу косинусна відстань між класами «кіт» та «собака» становить 0,9225, що показує високий рівень схожості й дозволяє заморозити більшу частину шарів моделі під час перенавчання. У базовому експерименті модель, навчена з нуля для розпізнавання котів, досягає максимальної валідаційної точності 0,769 за три епохи впродовж часу навчання 2951,4 секунд. Навчання моделі з нуля для класифікації собак дає дещо кращий результат — валідаційна точність підвищується до 0,803, а час навчання становить 2879,7 секунд. Натомість використання моделі, попередньо навченої на котах, із заморожуванням відповідної кількості шарів згідно з косинусною відстанню, забезпечує найкращий результат: валідаційна точність досягає 0,884, що істотно перевищує обидва випадки навчання з нуля. До того ж, час навчання зменшується до 2234,6 секунд, що свідчить про суттєву економію обчислювальних ресурсів.

Висновки

У роботі запропоновано метод динамічного трансферного навчання, що визначає оптимальну кількість шарів для заморожування на основі косинусної відстані між класами вихідного та цільового наборів даних. Проведені експерименти підтвердили, що такий підхід забезпечує підвищення якості моделі та скорочення часу навчання порівняно з традиційним навчанням з нуля. Використовуючи попередньо навчену модель та адаптивно зберігаючи релевантні ознаки нижчих рівнів, модель демонструє значне зростання точності, навіть за умов обмеженої кількості зображень, завдяки чому запропонований метод є ефективним інструментом для практичних застосувань. По-перше, адаптивне визначення кількості заморожених шарів дозволяє оптимально використовувати попередньо навчені шари, зберігаючи лише ті частини моделі, які справді узгоджуються з цільовим класом. Це забезпечує не лише значне скорочення часу тренування, але й стабільніше навчання. По-друге, експериментальні результати підтверджують, що динамічний підхід значно підвищує валідаційну точність у задачах з обмеженим обсягом даних. Отримане підвищення точності з 0,803 (навчання з нуля на собаках) до 0,884 у випадку трансферного навчання підтверджує, що метод не лише зменшує витрати часу, але й покращує якість моделі, завдяки чому він є ефективним інструментом для невеликих і середніх датасетів.

Метод має значний потенціал для масштабування на складніші набори даних і ширший діапазон прикладних задач. Оскільки косинусна відстань відображає семантичну близькість між класами, підхід може бути використаний для побудови адаптивних систем вибору оптимального рівня перенавчання для кожної категорії окремо, що особливо цінно у багатокласних або long-tail розподілах. У задачах детекції або сегментації, де моделі значно глибші, динамічне визначення шарів для заморожування може ще більше скоротити обчислювальні витрати без втрати точності. До того ж, метод потенційно може бути розширений до сценаріїв few-shot learning або continual learning, де важливо зберігати попередні знання та уникати катастрофічного забування. Таким чином, динамічний підхід створює основу для побудови гнучкіших, економічних і стійких моделей глибокого навчання, які адаптуються до нових задач на основі реальної семантичної близькості класів.

Отримані результати свідчать про те, що використання косинусної відстані як критерію для визначення глибини перенавчання є обґрунтованим і перспективним напрямом подальших досліджень. Цей підхід показав свою здатність адаптуватися до рівня семантичної схожості класів, що дозволяє раціональніше використовувати обчислювальні ресурси та уникати надмірного перенавчання. У майбутніх роботах доцільно розширити підхід на складніші архітектури та багатокласові сценарії, а також дослідити застосування методу в задачах детекції, сегментації та few-shot learning. Отже, запропонований метод формує основу для побудови гнучкіших, стійкіших і продуктивних систем глибокого навчання.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] S. J. Pan, and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010. <https://doi.org/10.1109/TKDE.2009.191>.
- [2] F. Zhuang, et al., "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, 2020. <https://doi.org/10.1109/JPROC.2020.3004555>.
- [3] A. Huang, "Similarity measures for text document clustering," *Proceedings of the Sixth New Zealand Computer Science Research Student Conference*, 2008, pp. 49-56.
- [4] J. Yosinski, et al., "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, vol. 27, pp. 3320-3328, 2014. <https://doi.org/10.48550/arXiv.1411.1792>.
- [5] D. A. Ivanov, "Reducing the training time of models using transfer learning," *Scientific Works of VNTU*, no. 3, 2024. <https://doi.org/10.31649/2307-5376-2024-3-25-30>.
- [6] S. Kornblith, J. Shlens, and Q. V. Le, "Do Better ImageNet Models Transfer Better?" *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2661-2671. <https://doi.org/10.1109/CVPR.2019.00277>.
- [7] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," *Advances in Neural Information Processing Systems*, vol. 32, 2019. <https://arxiv.org/abs/1902.07208>.
- [8] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "A Theory of Learning from Different Domains," *Machine Learning*, vol. 79, pp. 151-175, 2010. <https://doi.org/10.1007/s10994-009-5152-4>.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>.
- [10] A. Krizhevsky, *Learning Multiple Layers of Features from Tiny Images*, University of Toronto, 2009.

Рекомендована кафедрою комп'ютерних систем управління ВНТУ

Дата надходження: 22.12.2025
Дата прийняття до публікації: 27.03.2026
Дата публікації: 8.04.2026

Іванов Дмитро Анатолійович — аспірант кафедри інженерії програмного забезпечення, e-mail: www.ivanovda@gmail.com, <https://orcid.org/0000-0002-7386-4497>.

Державний університет «Житомирська політехніка», Житомир

D. A. Ivanov¹

Improving the Efficiency of Transfer Learning Through Dynamic Determination of the Number of Layers to Freeze Based on Class Similarity

¹Zhytomyr Polytechnic State University

In this research work, a method is presented that automatically determines the number of neural network layers to be frozen during transfer learning. The proposed approach is based on the use of cosine distance between the vector representations of classes from the source and target datasets, which makes it possible to evaluate their semantic similarity and accordingly regulate the depth of layer freezing for further fine-tuning. Unlike fixed strategies, the dynamic approach provides more flexible utilization of prior knowledge and reduces the need for excessive computations. The study employs the ResNet-50 architecture and class subsets from the CIFAR-10 dataset, for which mean feature vectors were generated and cosine distance was calculated. Based on these values, the method dynamically determined how many layers of the model should be frozen and kept unchanged, and which layers require retraining. The effectiveness of the method was evaluated by comparing classical training from scratch with training using this transfer learning approach. The obtained results demonstrate that the proposed method improves generalization quality and reduces training time, highlighting the advantages of dynamically determining the number of layers to freeze. The method can be applied in tasks where rapid adaptation, limited data, and efficient resource usage are essential. The proposed technique is promising, as it effectively combines model accuracy with reduced computational costs, enabling scalability, reusability of pretrained layers, and quick integration into various application domains, making it a valuable tool for future research and practical implementations.

Keywords: transfer learning, dynamic layer freezing, cosine distance, ResNet-50, CIFAR-10, resource reduction.

Ivanov Dmytro A. — Post-Graduate Student of the Chair of Software Engineering, e-mail: www.ivanovda@gmail.com, <https://orcid.org/0000-0002-7386-4497>